

Mariafilomena Anzalone

Guglielmo Tamburrini, *Etica delle macchine.*

Dilemmi morali per robotica e intelligenza artificiale,

Carocci, Roma 2020, pp. 151

“Il minuto era passato. Strana sensazione quando ti decidi e non riesci a deciderti. ‘Andare o no, andare o no?’ ripetevo ogni secondo quasi rabbrivendo”. Le parole con cui Dostoevskij, ne *L’Adolescente*, racconta l’esitazione, colma di trepidazione, vissuta dal giovane Arkadij di fronte alla possibilità di incontrare per la prima volta il suo fratellastro, descrivono un’esperienza piuttosto comune per gli esseri umani: decidere senza agire di conseguenza, tornare sui propri propositi, indugiare senza dare loro seguito, sebbene nulla di nuovo e significativo sia accaduto da quando sono stati formulati. Si tratta di un’esperienza che, invece, è del tutto estranea a quegli agenti macchinici che, combinando le potenzialità della robotica e dell’Intelligenza Artificiale (IA), sono in grado di scegliere un corso d’azione e portarlo a termine senza l’intervento da remoto di un operatore umano. Un drone armato, progettato per individuare e eliminare obiettivi militari, agirà di conseguenza, senza tentennamenti; lo stesso dicasi per un veicolo a guida autonoma (VA) programmato per percorrere la rete stradale. Una delle particolarità di questi dispositivi, come evidenzia l’interessante volume di Guglielmo Tamburrini, è data proprio dalla rapidità con cui passano dalla “decisione algoritmica” all’“azione robotica” (p. 11). Rapidità che se da un lato determina l’elevato livello di efficacia operativa di queste macchine, dall’altro riducendo fino ad annullarle le possibilità di un intervento umano che modifichi il corso d’azione intrapreso, ha delle rilevanti implicazioni di natura morale. L’Autore sceglie di esporle ed esplorarle all’interno di un confronto tra due tradizioni normative – l’etica dei doveri e l’etica delle conseguenze – che, soprattutto per quel che concerne l’etica dei doveri, sono presentate in una versione non priva di semplificazioni. L’Autore le assume, un po’ schematicamente, come punti di riferimento principali e contrapposti – sebbene non esclusivi – sia per individuare i problemi morali configurati dall’utilizzo di tali macchine che per articularne possibili soluzioni. All’interno di questa cornice, il volume offre una ricca e documentata ricognizione delle principali criticità etiche connesse alla crescente auto-

nomia operativa delle macchine, spaziando dalle questioni legate ai processi di trasformazione del lavoro, delle professioni altamente qualificate (come quella del chirurgo) e delle responsabilità ad esse collegate, alle implicazioni derivanti dai sistemi di sorveglianza e controllo sociale resi possibili da tali tecnologie.

Questa panoramica di carattere generale, a cui è dedicato l'ultimo capitolo del volume, è preceduta dall'esame di alcuni casi di studio – riguardanti, nello specifico, i veicoli a guida autonoma e le armi autonome – che, con grande efficacia narrativa, consentono al lettore di entrare nel vivo e nel merito dei problemi suscitati da tali macchine. Problemi che, come evidenziato anche nel sottotitolo del volume, assumono sovente la forma di veri e propri dilemmi morali.

Esemplare, in tal senso, è l'analisi del caso delle collisioni inevitabili dei VA che, ricalcando il famoso *trolley problem* elaborato da Philippa Foot, si configura quando il VA non può evitare di impattare contro un ostacolo. A parità di condizioni concomitanti (velocità, angolo d'impatto, etc.), il VA potrebbe trovarsi nella situazione di dover scegliere, ad esempio, se collidere con un ciclista munito di casco o con uno che ne è privo. In nome del paradigma consequenzialista cui si ispira la stessa idea di sviluppare veicoli del genere – la cui introduzione su larga scala mira ad aumentare la sicurezza stradale minimizzando imparzialmente i danni attesi per le persone – si potrebbe ritenere giusto collidere con il ciclista munito di casco che, essendo più protetto, corre meno rischi di riportare gravi conseguenze. Ma, se nell'immediato una scelta del genere sembra limitare i danni, a lungo termine essa rischierebbe di condurre in una direzione completamente opposta: consapevoli dei criteri con cui vengono programmati i VA, i ciclisti potrebbero decidere di non indossare più il casco, visto che li renderebbe bersagli privilegiati in caso di collisioni inevitabili. Con l'effetto, del tutto contrario rispetto a quello perseguito dai VA, di accrescere i pericoli per l'integrità fisica delle persone.

Se anche immaginassimo di superare il problema dei ciclisti facendo circolare i VA su percorsi stradali a loro preclusi, non potremmo comunque escludere a priori scenari di collisione inevitabile in cui a rischio è la vita del passeggero. In questi casi, secondo quali finalità andrà programmato il VA? Dovrà privilegiare sistematicamente la vita del passeggero o calcolare imparzialmente i danni attesi per tutte le persone coinvolte nell'impatto? È evidente che una scelta del genere ha delle implicazioni notevoli sia sul piano individuale che dell'accettabilità sociale: quanti di noi acquisterebbero un VA e lo utilizzerebbero per portare a scuola i propri figli sapendo che è programmato per non anteporre la loro protezione a quella degli altri individui? Ma, allo stesso tempo, da pedoni, ci sentiremmo sicuri nell'attraversare strade in cui circolano VA programmati per privilegiare ad ogni costo i loro passeggeri?

Come l'Autore non manca di evidenziare, casi del genere portano alla luce un elemento che differenzia e complica il quadro rispetto alla tradizionale formulazione dei dilemmi del *trolley*: qui non si tratta di effettuare una scelta *una tantum*, "in una situazione che non prevede ripetizioni" (p. 26), ma di programmare, *ex ante*, una macchina perché decida sistematicamente di intraprendere, date certe condizioni, un determinato corso di azione. La programmata reiterazione delle scelte impone, a sua volta, di ampliare di molto l'orizzonte temporale, valutando a lungo termine il loro impatto in una prospettiva che non può ignorarne le ricadute di carattere sociale e politico. Di qui la necessità, fortemente sottolineata dall'Autore, di avviare un confronto pubblico tra tutti i soggetti interessati, confronto in cui alla filosofia morale spetta il compito di contribuire in modo decisivo all'elaborazione di "politiche etiche accettabili e socialmente condivise" (p. 31). La loro esigenza si fa ancora più stringente quando si considera il caso delle armi autonome il cui sviluppo e utilizzo è già di per sé fortemente controverso, data la natura esplicitamente offensiva di tali dispositivi progettati per individuare e colpire obiettivi militari. L'Autore dedica ampio spazio agli argomenti contro le armi autonome riconducibili all'etica delle conseguenze e, soprattutto, all'etica dei doveri, delineando un ventaglio di criticità che non sono risolvibili solo mediante un progressivo raffinarsi delle tecnologie disponibili. Anche se, in un futuro non troppo lontano, le armi autonome potessero assicurare una selezione corretta e precisa degli obiettivi militari, resterebbero comunque aperte una serie di questioni connesse al rispetto dei diritti e della dignità personale di chi è coinvolto in un conflitto bellico, ai potenziali effetti di destabilizzazione politica derivanti dalla loro utilizzazione e, più in generale, alla responsabilità morale e giuridica di scelte che mettono a repentaglio la vita e l'integrità fisica delle persone.

In un quadro che, anche dal punto di vista giuridico, appare ancora fortemente inadeguato, il dibattito pubblico è attraversato da richieste di moratoria o di totale messa al bando delle armi autonome. Ma, come l'Autore opportunamente rileva – lasciando emergere il costante rinvio tra le problematiche etiche di natura applicativa e il piano più squisitamente teorico – una discussione del genere necessita di una preliminare riflessione su cosa significhi parlare di autonomia in riferimento a tali macchine.

Ed è questo il nodo gordiano a cui finiscono per essere ricondotti i vari e variegati problemi che il volume affronta e discute. L'Autore prova a scioglierlo delimitando innanzitutto il perimetro all'interno del quale si colloca l'autonomia esplicitamente operativa di queste macchine, che sono autonome dagli agenti umani ma non dal collegamento e dall'iterazione con altri agenti artificiali. Esso può essere circoscritto dalle condizioni del "per-da-dove": "autonomia operativa *per* un dato compito,

svolto indipendentemente *da* agenti di un certo tipo, *laddove* il contesto operativo sia congeniale al regolare funzionamento della macchina” (p. 59). Ma condizioni del genere, soddisfatte anche dai termostati deputati al controllo della temperatura nelle nostre case, non sono sufficienti a descrivere l'autonomia di dispositivi che sfruttano le più avanzate risorse della robotica e dell'IA. Non solo perché essi sono in grado di svolgere compiti la cui esecuzione, per gli uomini, implica notevoli assunzioni di responsabilità morale; ma anche e soprattutto perché la loro indipendenza dal controllo umano si salda alla capacità di apprendere dall'esperienza, di adattarsi e rispondere in modo flessibile quanto efficace ai cambiamenti del contesto operativo. I noti progressi dei sistemi di IA nel campo del *machine learning* e del *deep learning* consentono loro di processare e vagliare in modo pressoché istantaneo enormi quantità di dati, di selezionare, grazie a reti neurali artificiali, i casi più rilevanti per il raggiungimento degli obiettivi, ma anche di imparare a svolgere compiti per i quali non erano stati esplicitamente programmati. La sequenza di operazioni e processi che tali sistemi mettono in atto può essere, in alcuni casi, talmente raffinata e complessa da risultare poco o per nulla decifrabile per l'uomo che non è messo nelle condizioni di valutare fino in fondo “la bontà di una scelta automatica” (p. 110).

Ma è proprio sulla linea di confine che traduce l'autonomia di queste macchine in un automatismo che sfugge alla piena comprensione dell'uomo che l'Autore individua il discrimine tra l'autonomia umana e quella meccanica: quest'ultima non è consapevole di sé, delle proprie intenzioni, né è capace di avvertire le scelte che effettua come qualcosa di proprio. Anche se applica delle regole di natura morale, non è in grado di spiegare il perché tali regole abbiano determinato la scelta di un corso d'azione piuttosto che di un altro. Mancando attualmente di capacità introspettive, le macchine non sono capaci di fornire motivazioni e quindi, secondo l'Autore, non possono essere considerate dotate di un'autentica autonomia di natura morale. Le loro scelte morali, pertanto, sono scelte “in cerca d'autore” (p. 137), di un autore che, almeno al momento, non può che essere umano.

Da questo punto di vista, il volume ha l'indubbio merito di mostrare, con grande chiarezza e efficacia argomentativa, che una riflessione etica sull'autonomia delle macchine non corrisponde *tout court* ad un'indagine sulla fatale perdita di autonomia dell'uomo, destinato a soccombere di fronte al potere smisurato di sofisticati agenti meccanici. Essa, piuttosto, trova il suo cardine proprio nell'autonomia dell'essere umano in quanto agente morale, negli scopi che intende perseguire utilizzando tali dispositivi e nei valori morali che, attraverso di essi, intende realizzare. Non dobbiamo dimenticare, infatti, che, se tali macchine vanno ben oltre l'applicazione di regole generali a casi particolari e riescono a formulare

esse stesse regole di comportamento, lo fanno sempre all'interno di un orizzonte di fini fissati, stabiliti e continuamente riconfigurati dall'uomo. Del resto, l'autonomia dell'uomo, anche se può dar luogo ad automatismi, non coincide mai fino in fondo con essi né si lascia misurare solo in termini di velocità deliberativa ed esecutiva. Perché, per il giovane Arkadij, così come per tutti gli esseri umani, il tempo delle scelte è sempre e comunque anche il tempo delle esitazioni, dei dubbi e degli opportuni ripensamenti.